

大数据释义

第一堂课，由陈老师讲授大数据释义。



大数据释义：

1980 年，未来学家阿尔文·托夫勒便在《第三次浪潮》一书中，将大数据热情地赞颂为“第三次浪潮的华彩乐章”。维基百科引到大数据是指庞大到了无法通过人工在合理的时间内进行处理的数据。

本堂课陈老师对大数据进行了精心的解读。

“数”指的是人们计数、标记、度量事物的单位，是人类精确描述事物的基础；

“据”本意是倚靠，《说文》中提到“据，杖持也”，后演化为事物依靠的证据，成为分析事物的标准；

“大”指规模、形态、数量、程度等方面超过一般或超过所比较的对象，《易》中释文有云“大有，包容丰富之象。”大数据中的“大”则指表达了数据的数量多，涉及领域宽泛，数据结构多样等等诸多特点。

90 年代的启蒙：



陈老师回忆了他所经历过的计算机和大数据发展史。90年代初期，实验室用 15,000 的高价买进了一台计算机。大家围坐在计算机面前，谁都不敢上前敲击。信息化的狂潮悄然而至，人们窥探科技时代的影子，却难以洞悉电子世界的秘密。90年代中期，互联网，计算机领域犹如即将打开新世纪大门的一把金钥匙，计算机专家仅仅花费一点时间搭建小网站就可以在在市场上获得一笔可观的收入。

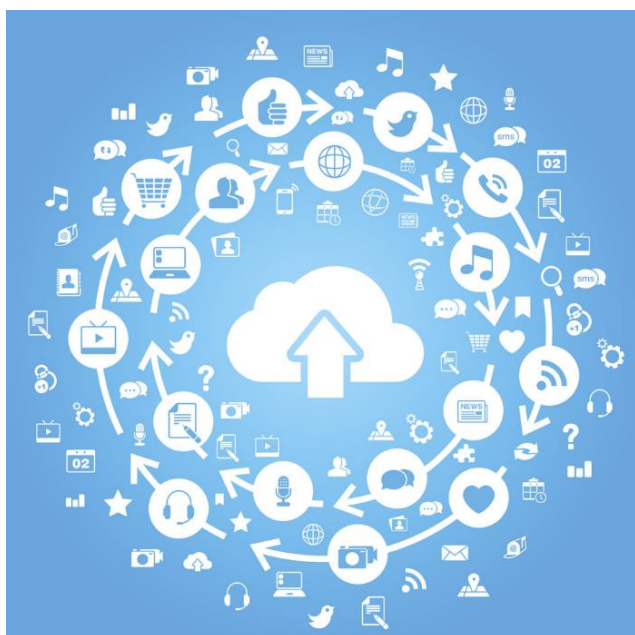


90年代末，为了让医护人员对人体有全面而详细的认识，同时让从医人员快速掌握人体各部位的结构，一种被称为虚拟人的三维重建人体诞生了。人体切片数据集的构建方法十分原始，虚拟人拥有三维人体详细结构，做法是利用尸体冰冻后再精确切割。这种做法在当时也引起了很大的争议。数据收集的困难成为了大数据研究领域的瓶颈。

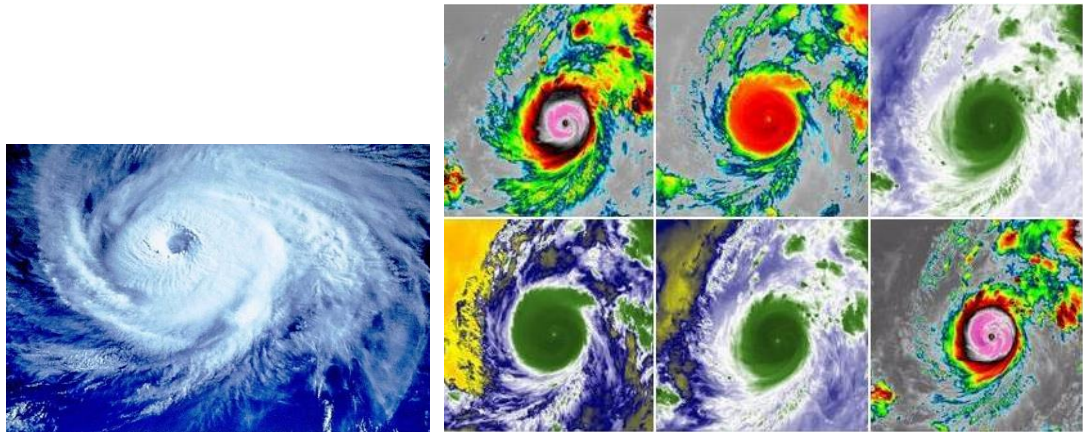
21 世纪初的探索



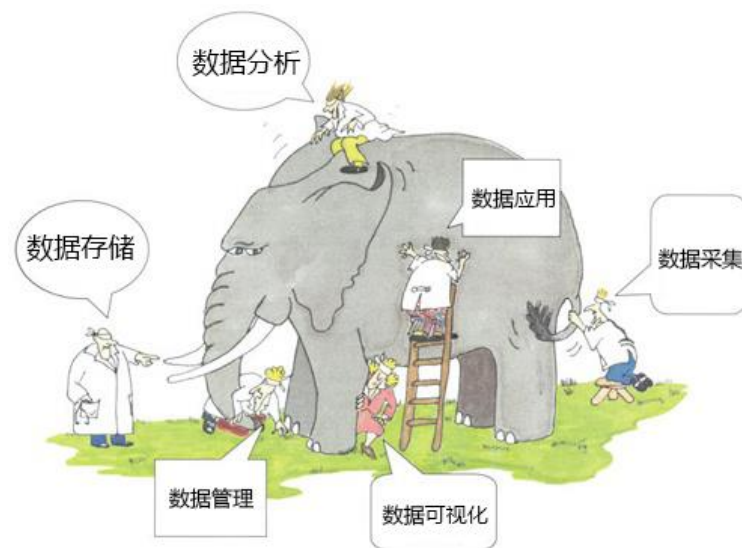
20 世纪人们还无法对大数据进行精确的解读，直到 2000 年，随着计算机的普及和各类数据收集工具的诞生。规模难以想象的海量数据涌入了各行各业的大门。大数据的概念开始被大众所接受。“大数据是长时间的积累”陈老师说，“信息是流动的知识，知识是沉淀的信息。”



计算机的发展遵循着摩尔定律，大数据的研究也瞬息万变。2005 年，数据到了云端，计算处理大量数据的需求在云端得以实现，也同时拉动了各个领域的专家对大数据进行研究探索。

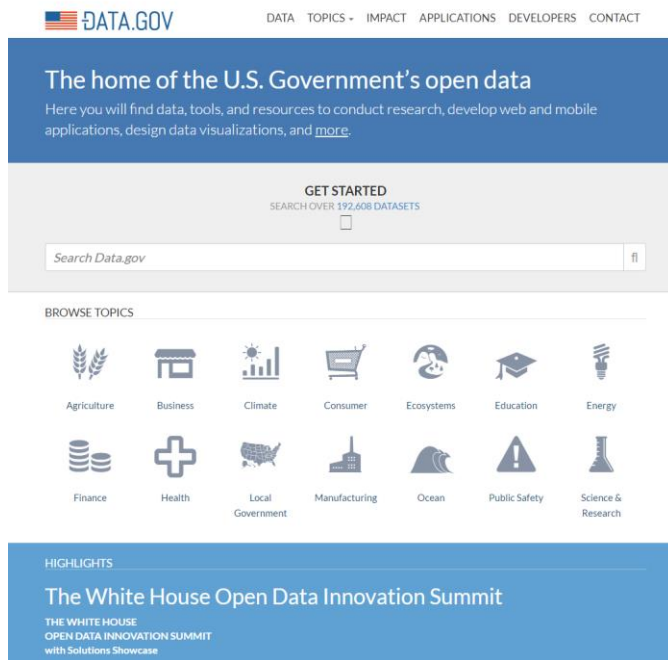


例如气象领域，对于台风眼的研究对于台风的防范和预测起到关键的作用，各种可视分析的方法应用大数据使得台风预测成为可能。

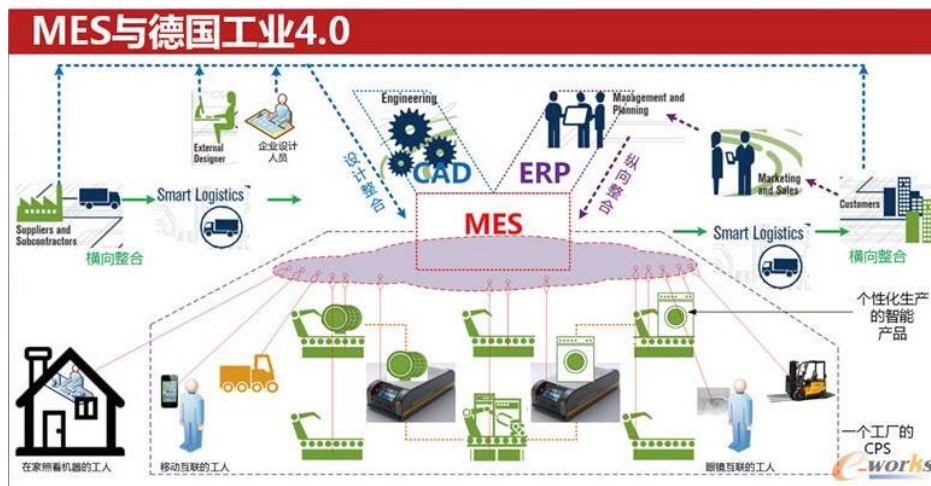


然而大数据领域的研究就像盲人摸象，不同领域的研究人员专注于自己的领域来探索大数据。分散的研究目的，独立的任务流程，不同的探索方向都使得人们对大数据的认识犹如管中窥豹，只能可见一斑。

数据的绽放



“大数据是一个理念”陈老师提到“先有数据才有资源，因此要有存储数据的理念”。2009年，得到奥巴马的许可，美国政府的首任首席信息官昆德拉建立了全面开放政府拥有的公共数据的网站 data.gov。交通，工业，税收，研究等领域的数据全都包罗其中。这是一个信号，提醒着人们研究大数据，需要领域之间的合作。



近几年不同的技术不断的整合起来。CPS 是工业界改革的核心。德国的工业 4.0 和 MES 强调了制造业+众包+数据化、智慧化的生产模式。



“物联网是五官，云计算是血液，大数据是大脑。”人的认知+信息世界（存储数据）是新时代各行各业的发展关键。GoldCorp 公司公开了整个金矿矿区的地质资料，让各行各业的人们分析选取开矿地点，最终实现了公司巨额的盈利。

信息提供了解决问题的能力，然而信息只是这种 CPS 系统中的一个方面，不只是信息通讯、采集、管理等内容，更有智能控制、自律制造等成分。



云栖大会上，曾为浙江大学教授，现任阿里巴巴集团首席技术官王坚博士说道“世界上最遥远的距离是红绿灯跟交通监控摄像头的距离，它们都在一根杆子上，但是从来就没有通过数据被连接过。中国有的大城市有将近 60 万个摄像头，但数据得不到利用，因为如果不借助人工智能，需要 120 万人才能在当天把摄像头的数据看完。杭州数据大脑第一次让摄像头的数据能够用来指挥交通信号灯，而交通治理只是个开始，更重要的是数据开始为社会产生价值。”

结语：

当计算佩戴上飞速的轮滑，当大数据建成会思考大厦，当人机交互成为解读世界的密码，你觉得，智慧的生活还会远吗？